



Association analysis of genetic variants with type 2 diabetes in a mongolian population in China

Bai, Haihua; Liu, Haiping; Suyalatu, Suyalatu; Guo, Xiaosen; Chu, Shandan; Chen, Ying; Lan, Tianming; Borjigin, Burenbatu; Orlov, Yuriy L.; Posukh, Olga L.; Yang, Xiuqin; Guilan, Guilan; Osipova, Ludmila P.; Wu, Qizhu; Narisu, Narisu

Published in:
Journal of Diabetes Research

DOI:
[10.1155/2015/613236](https://doi.org/10.1155/2015/613236)

Publication date:
2015

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Bai, H., Liu, H., Suyalatu, S., Guo, X., Chu, S., Chen, Y., Lan, T., Borjigin, B., Orlov, Y. L., Posukh, O. L., Yang, X., Guilan, G., Osipova, L. P., Wu, Q., & Narisu, N. (2015). Association analysis of genetic variants with type 2 diabetes in a mongolian population in China. *Journal of Diabetes Research*, 2015, [613236].
<https://doi.org/10.1155/2015/613236>

Research Article

Association Analysis of Genetic Variants with Type 2 Diabetes in a Mongolian Population in China

Haihua Bai,¹ Haiping Liu,¹ Suyalatu Suyalatu,¹ Xiaosen Guo,^{2,3} Shandan Chu,¹ Ying Chen,² Tianming Lan,² Burenbatu Borjigin,¹ Yuriy L. Orlov,^{4,5} Olga L. Posukh,^{4,5} Xiuqin Yang,⁶ Guilan Guilan,¹ Ludmila P. Osipova,⁴ Qizhu Wu,¹ and Narisu Narisu⁷

¹Inner Mongolia University for the Nationalities, Tongliao, Inner Mongolia 028000, China

²BGI-Shenzhen, Shenzhen 518083, China

³Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark

⁴The Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (SB RAS), Novosibirsk 630090, Russia

⁵Novosibirsk State University, Novosibirsk 630090, Russia

⁶College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China

⁷Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

Correspondence should be addressed to Qizhu Wu; qizhu_wu2014@163.com and Narisu Narisu; narisu@mail.nih.gov

Received 7 February 2015; Revised 22 June 2015; Accepted 24 June 2015

Academic Editor: Andrea Tura

Copyright © 2015 Haihua Bai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The large scale genome wide association studies (GWAS) have identified approximately 80 single nucleotide polymorphisms (SNPs) conferring susceptibility to type 2 diabetes (T2D). However, most of these loci have not been replicated in diverse populations and much genetic heterogeneity has been observed across ethnic groups. We tested 28 SNPs previously found to be associated with T2D by GWAS in a Mongolian sample of Northern China (497 diagnosed with T2D and 469 controls) for association with T2D and diabetes related quantitative traits. We replicated T2D association of 11 SNPs, namely, rs7578326 (*IRS1*), rs1531343 (*HMGA2*), rs8042680 (*PRC1*), rs7578597 (*THADA*), rs1333051 (*CDKN2*), rs6723108 (*TMEM163*), rs163182 and rs2237897 (*KCNQ1*), rs1387153 (*MTNR1B*), rs243021 (*BCL11A*), and rs10229583 (*PAX4*) in our sample. Further, we showed that risk allele of the strongest T2D associated SNP in our sample, rs757832 (*IRS1*), is associated with increased level of TG. We observed substantial difference of T2D risk allele frequency between the Mongolian sample and the 1000G Caucasian sample for a few SNPs, including rs6723108 (*TMEM163*) whose risk allele reaches near fixation in the Mongolian sample. Further study of genetic architecture of these variants in susceptibility of T2D is needed to understand the role of these variants in heterogeneous populations.

1. Introduction

Type 2 diabetes (T2D) is a complex disease characterized by insulin resistance and pancreatic beta-cell dysfunction. In China, 9.7% and 15.5% of the entire population suffer from T2D and prediabetes, respectively [1]. Given recent advances in genotyping and sequencing technology, the GWAS have contributed significantly to the identification of susceptibility loci for T2D and many other complex disorders. At least 80 loci conferring susceptibility to T2D have been identified to date [2, 3] (<http://www.genome.gov/gwastudies/>) and

the genetic architecture underlying T2D varies substantially between populations of different ethnic backgrounds [4, 5]. Studying genetics of T2D in multiethnic cohorts has been insightful for fine-mapping causal variants and identifying new loci [6], demonstrating the use of investigating common variants in different ethnic samples.

There are 10 million Mongolians currently living in various regions of Asia [7]. To our knowledge, there is only one T2D association study conducted in a Mongolian sample, which replicated association of variants in *KCNQ1* and *ABCC8* with T2D [8]. However, the small-sized study (177

cases and 216 controls) has low power to detect variants with smaller effect. The prevalence of T2D in Mongolians in China has grown among the adult urban population from 1.86% in 1980 to 5.6% in 2012 [9–11]. In our study, we aim to explore genetic risks of 34 T2D SNPs previously reported by GWAS in a larger Mongolian sample. Twenty-eight SNPs passed rigorous quality control filtering and 11 of them showed significant association with T2D (Bonferroni corrected $P < 0.05$). The SNP with the strongest T2D association, rs7578326, has the risk allele significantly associated with increased levels of TG. We demonstrated the need of further study of allelic difference of T2D associated SNPs in diverse populations.

2. Methods and Materials

2.1. Ethics Statement. This study was approved by the Institutional Review Board of the Affiliated Hospital of Inner Mongolia University for the Nationalities and complied with the Declaration of Helsinki. The written informed consent was obtained from each participant.

2.2. Study Population. We collected whole blood samples from 986 individuals of Mongolian ethnicity from Inner Mongolia, China. The sample was comprised of 511 T2D cases and 475 healthy normoglycemic controls, of which 497 cases and 469 controls passed quality control filtering and were used for subsequent analysis (see below). Cases were registered based on the World Health Organization (WHO) criteria [12] of fasting plasma glucose concentration ≥ 7 mmol/L or 2-h plasma glucose concentration ≥ 11.1 mmol/L and were admitted to the affiliated hospital of the Inner Mongolia University for Nationalities. Nondiabetic healthy controls were selected based on matching sex and ethnic background from the same region. Aside from the diagnosis of T2D, we collected other diabetes related lipid traits, such as TC, HDL-C, LDL-C, and TG, for each individual. We collected certain life style information (smoking and drinking habits), waist circumference (WC), and body mass index (BMI) of each participant as well.

2.3. Selection of SNPs and Genotyping. We selected a list of SNPs previously found to be associated with T2D based on the NHGRI GWAS catalog [2] (available at <http://www.genome.gov/gwastudies/>, November, 2012). Candidate SNPs were initially selected with the following considerations: (1) SNPs found to be associated with T2D in an Asian sample were given higher priority (rs6723108 and rs5945326 were added after the initial selection date); and (2) subsequently SNPs found to be associated with multiple studies were included. We were able to genotype 34 SNPs located in or near 33 candidate genes (see Supplementary Material available online at <http://dx.doi.org/10.1155/2015/613236>). We included two SNPs around *KCNQ1*, because those have been reported to be associated with Asian populations [13, 14].

We estimated the concentration of isolated genomic DNA using Qubit dsDNA BR Assay Kit (Invitrogen, USA), and the DNA solution was further diluted to a concentration of 10 ng/ μ L. We designed the targeted sequencing primers

and redesigned the primer sets with dispersed or weak electrophoretic bands. To prepare the chip array, we used a multisample nanodispenser (WaferGen, USA) to disperse DNA and primers into SmartChip MyDesign Chip (WaferGen, USA). Following the polymerase chain reaction (PCR) amplification, we purified PCR products through Agencourt AMPure XP-medium beads to get mixed Illumina pair-end libraries. Insert sizes were calculated by Agilent 2100 bio-analyzer (Agilent, USA) and concentrations were estimated by Real Time PCR. Sequencing was performed on either Illumina MiSeq or on Illumina HiSeq 2500. All sequencing steps were in strict accordance with Illumina recommended protocols.

The final sequencing depth reached $> 200\times$, and the length of pair-end reads was 100 bp. Reads with an average base quality of ≥ 20 were kept for further analysis. BWA [15] (v0.5.9, available at <http://bio-bwa.sourceforge.net/>) was used to map all clean reads against the human reference genome of hg19 allowing ≤ 3 mismatches across a single read. Samtools mpileup (v0.1.18, available at <http://samtools.sourceforge.net/>) command was used to obtain SNP genotypes as described in [16].

These genotypes were further filtered according to the following criteria: SNPs with $\geq 5\%$ of missing call rate across the samples. Samples with $\geq 3\%$ of missing genotypes (which corresponds to 10% of missing SNP call rate) were removed. We tested SNPs for Hardy-Weinberg Equilibrium (HWE) and excluded SNPs with HWE P value $< 1 \times 10^{-6}$ in unaffected individuals. Twenty-eight SNPs of 966 samples (497 cases and 469 controls) passed the quality control filtering, and the overall genotype call rate is 99.3% or higher across the sample.

2.4. Statistical Analysis. We tested association between candidate SNPs and the status of T2D using logistic regression (likelihood ratio test) by adjusting for the effects of age, sex, and BMI. The study-wide significance was determined by applying Bonferroni correction using 28 tested SNPs (P value $\leq 0.05/28 = 1.8 \times 10^{-3}$). We tested association with diabetes related quantitative traits (TC, HDL-C, LDL-C, and TG) across both T2D cases and controls using linear regression with the age, sex, BMI, and T2D status as covariates. All quantitative trait measures were normalized by quantile normalization and the normalized values were used in the analyses. Formal statistical tests, including 95% confidence intervals (CI), were performed using EPACTS [17] (v3.2.6, available at <http://www.sph.umich.edu/csg/kang/epacts/>). Differences in population structure between the Mongolian sample (healthy controls) and healthy Caucasian (CEU) or Chinese (CHB and CHS) samples of 1000 G project [18] (<http://www.1000genomes.org/>) were estimated by comparing risk allele frequency and the Wright's fixation index (F_{ST}) using plink [19, 20]. Comparison of trait values between cases and controls was conducted using Student's t -test.

3. Results

After rigorous sample and marker level quality control filtering, genotypes of 28 SNPs on 966 individuals (including

TABLE 1: Clinical characteristics of study population.

Traits	Controls ^a	Cases	<i>P</i> value for normalized trait value ^b
Samples (<i>N</i>)	469	497	
Sex (male/female)	168/301	214/283	
Smoking (%)	22.2	31.6	
Drinking (%)	38.6	27.2	
Age (years)	40.5 ± 15.8	54.3 ± 10.1	
BMI (kg/m ²)	24.0 ± 4.8	26.1 ± 4.0	<0.001
WC (cm)	81.7 ± 14.5	92.0 ± 11.6	<0.001
TC (mmol/L)	4.6 ± 1.2	5.4 ± 1.2	<0.001
TG (mmol/L)	1.9 ± 1.5	3.1 ± 2.9	<0.001
HDL-C (mmol/L)	1.3 ± 0.3	1.4 ± 1.0	0.775
LDL-C (mmol/L)	2.7 ± 0.9	3.2 ± 1.0	<0.001

^a Mean ± SD.^b Student's *t*-test. BMI: body mass index; WC: waist circumference; TC: total cholesterol; TG: triglyceride; HDL-C: high density lipoprotein cholesterol; LDL-C: low density lipoprotein cholesterol.

497 with T2D cases and 469 nondiabetic ethnically matched controls) were kept for subsequent analyses. Clinical characteristics of the sample are summarized in Table 1. Overall, consistent with previous studies [21], T2D cases in current study have higher TC, TG, and LDL-C values compared to controls and have comparable HDL-C values with the controls, indicating that TC, TG, and LDL-C are among risk factors for T2D in the Mongolian sample.

Table 2 presents the association results between the 28 SNPs and T2D status. Of the 28 SNPs tested, 11 SNPs were significantly associated after correcting for multiple testing ($P < 1.8 \times 10^{-3}$). We replicated a T2D association near *KCNQ1* (rs2237897; OR = 1.39; $P = 0.002$), originally identified in a Japanese population [14], and subsequently replicated in another Mongolian population sample with the same ethnic background as our sample [22]. We also replicated T2D association of three SNPs initially identified in Asian samples, namely, rs163182 (*KCNQ1*) [13] in Japanese, rs6723108 (*TMEM163*) [23] in Indians, and rs10229583 (*PAX4*) [24] in Chinese samples. In addition, we replicated associations of seven T2D SNPs previously identified in European populations. To our knowledge, the association of rs7578326 (*IRS1*), rs1531343 (*HMG2*), rs8042680 (*PRC1*), rs1387153 (*MTNR1B*), rs7578597 (*THADA*) [25], rs243021 (*BCL11A*) [26], and rs1333051 (*CDKN2*) [27] was for the first time replicated in an Asian sample.

Among the four lipid related traits tested, we only observed a single significant association of T2D risk allele A of rs7578326 (*IRS1*) with TG level ($P = 0.0004$, OR = 3.4, and CI 1.7–6.6). Mean TG level for AA, AG, and GG is 2.65, 2.12, and 2.03 mmol/L (Figure 1). Individuals homozygous for the risk allele A have 25% higher TG values compared to heterozygotes. Heterozygotes have 4% higher TG compared to those homozygous for the nonrisk allele G.

Most of the SNPs selected for testing in our sample have been identified to be associated with T2D initially in

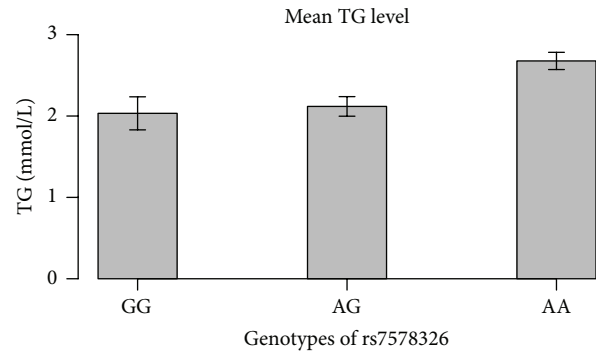


FIGURE 1: Correlation of mean triglyceride level with the genotypes of rs7578326 in the Mongolian sample. *y*-axis represents TG (mmol/L) level, and *x*-axis represents samples with different genotypes. Error bar represents standard error of TG in the samples with the specific rs7578326 genotype. Number of samples with genotypes GG, AG, and AA are 47, 248, and 647, respectively, across all samples including T2D cases and controls.

European populations. The lack of replication of 17 T2D loci identified by the GWAS studies in other populations prompted us to look at the differences of allelic architecture in the SNPs tested for the T2D association in our population. We observed variable risk allele frequency difference between our sample and 1000 G Caucasian (CEU) and Chinese populations (Table 3). Six out of 11 SNPs that have risk allele frequency 10% higher in the Mongolian sample compared to the Caucasian sample in 1000 G panel showed significant association with T2D in our study, and only two out of nine SNPs that have 10% or lower risk allele frequency compared to CEU showed significant association with T2D. Although statistically not significant, this observation shows a trend of overabundance of T2D associated SNPs in those with high frequency of risk alleles in the Mongolian population. In addition, we calculated Wright's fixation index to estimate whether the heterogeneity of a tested SNP is different between the Mongolian sample and Caucasian or Chinese samples. Noteworthy, a T2D associated SNP, rs6723108 (*TMEM163*), has reached near fixed high risk allele frequency (0.98 in our sample compared to 0.51 in the Caucasian sample of 1000 G project) and has a substantial population difference with the Caucasian (F_{ST} as high as 0.61). This trend is also present for the case of rs8042680 (*PRC1*), which has much higher risk allele frequency in the Mongolian sample compared to the Caucasians (0.92 versus 0.72; $F_{ST} = 0.67$). Although it is difficult to postulate the cause of the population difference, high proportion of Mongolians appear to carry this risk allele.

4. Discussion

In this study, we chose to examine the association of 34 GWAS SNPs, previously identified in European and East Asian populations, with susceptibility to T2D in a Mongolian sample from China. Six SNPs did not pass quality control filtering and were excluded from the analysis.

TABLE 2: Association of 28 SNPs tested with T2D in a Mongolian sample.

Chromosome	Nearby gene	SNP	Risk allele	Risk allele frequency		<i>P</i> value	OR	OR L95 ^a	OR H95 ^b
				Cases	Controls				
2	<i>IRS1</i>	rs7578326	A	0.843	0.792	6.4E – 08	1.27	0.98	1.64
12	<i>HMGA2</i>	rs1531343	C	0.111	0.106	6.2E – 07	1.08	0.79	1.49
15	<i>PRCI</i>	rs8042680	A	0.953	0.916	5.6E – 06	1.32	0.86	2.00
2	<i>THADA</i>	rs7578597	T	0.989	0.980	8.0E – 06	1.92	0.76	4.76
9	<i>CDKN2</i>	rs1333051	A	0.836	0.833	7.3E – 05	1.05	0.80	1.39
2	<i>TMEM163</i>	rs6723108	T	0.997	0.980	7.9E – 05	7.69	2.00	33.33
11	<i>KCNQ1</i>	rs2237897	C	0.676	0.585	8.7E – 05	1.39	1.11	1.72
11	<i>KCNQ1</i>	rs163182	C	0.427	0.358	1.0E – 04	1.37	1.11	1.69
11	<i>MTNR1B</i>	rs1387153	T	0.420	0.358	2.7E – 04	1.17	0.95	1.44
2	<i>BCL11A</i>	rs243021	A	0.715	0.696	1.2E – 04	1.33	1.06	1.67
7	<i>PAX4</i>	rs10229583	G	0.848	0.817	4.5E – 04	1.28	0.97	1.69
3	<i>IGF2BP2</i>	rs4402960	T	0.320	0.274	5.0E – 02	1.21	0.97	1.50
2	<i>GRB14</i>	rs3923113	A	0.847	0.812	1.7E – 02	1.20	0.93	1.56
13	<i>SPRY2</i>	rs1359790	G	0.698	0.679	9.9E – 03	1.18	0.95	1.47
3	<i>ADAMTS9</i>	rs4607103	C	0.594	0.566	1.2E – 01	1.11	0.88	1.39
11	<i>KCNJ11</i>	rs5215	C	0.370	0.336	6.1E – 03	1.03	0.84	1.25
10	<i>HHEX-IDE</i>	rs5015480	C	0.223	0.219	5.7E – 03	1.01	0.80	1.29
10	<i>TCF7L2</i>	rs7903146	T	0.053	0.061	1.1E – 01	1.01	0.66	1.57
11	<i>CENTD2</i>	rs1552224	A	0.896	0.891	5.3E – 02	1.00	0.72	1.39
15	<i>RASGRP1</i>	rs7403531	T	0.375	0.359	2.5E – 02	0.98	0.80	1.20
4	<i>WFS1</i>	rs1801214	T	0.947	0.943	5.8E – 02	0.98	0.62	1.54
16	<i>FTO</i>	rs8050136	A	0.162	0.166	1.3E – 01	0.98	0.75	1.30
8	<i>TP53INP1</i>	rs896854	T	0.351	0.389	3.1E – 02	0.93	0.76	1.15
12	<i>LGR5</i>	rs7961581	C	0.219	0.239	6.1E – 02	0.92	0.72	1.17
10	<i>GRK5</i>	rs10886471	C	0.755	0.774	2.2E – 02	0.86	0.67	1.10
6	<i>C6orf57</i>	rs1048886	G	0.148	0.212	1.2E – 04	0.67	0.51	0.89
15	<i>ZFAND6</i>	rs11634397	G	0.178	0.231	2.6E – 05	0.67	0.52	0.86
5	<i>ZBED3</i>	rs4457053	G	0.055	0.090	1.4E – 02	0.63	0.43	0.94

^aLower boundary of 95% confidence interval (CI) of odd's ratio.^bUpper boundary of 95% CI of odd's ratio.

This study confirmed the T2D association of rs2237897 in *KCNQ1* that were reported in European, Mexican, Chinese, Japanese, and Mongolian populations [8]. We also replicated the T2D association of three SNPs previously identified in Asian samples, namely, rs163182 (*KCNQ1*) [13] in Japanese, rs6723108 (*TMEM163*) [23] in Indians, and rs10229583 (*PAX4*) [24] in Chinese samples. In addition, our study replicated T2D association of seven additional GWAS SNPs in our sample. To our knowledge, the T2D association of rs7578326 (*IRS1*), rs1531343 (*HMGA2*), rs8042680 (*PRCI*), rs1387153 (*MTNR1B*), rs7578597 (*THADA*) [25], rs243021 (*BCL11A*) [26], and rs1333051 (*CDKN2*) [27] was replicated in an East Asian sample for the first time. This indicates that our study was able to replicate the result obtained in a population with the similar ethnic background and extended the replication of several other loci in an Asian population.

Although the association is not significant after the multiple test correction, additional six SNPs show the

OR trend consistent with the original reported studies. SNPs found to be associated with T2D in Asian populations, rs1048886 (*C6orf57*) [5], rs4402960 (*IGF2BP2*) [28], rs5015480 (*HHEX, IDE*), rs1359790 (*SPRY2*) [29], rs1552224 (*CENTD2*), rs3923113 (*GRB14*), rs5215 (*KCNJ11*), rs7903146 (*TCF7L2*) [6], rs10886471 (*GRK5*), and rs7403531 (*RASGRP1*) [30], were not replicated in our study. This observation clearly suggests the following: (1) our study has a limited power because of relatively small sample size, which warrants further confirmation of these SNPs in a larger sample from the same ethnic group; (2) our sample has variable degrees of difference in risk allele frequency compared to Caucasian population where most original GWAS were conducted, and (3) the selection of control individuals was not matched precisely with the cases, in particular, with respect to age, and there is a trend that the healthy controls are younger compared to the cases. However, we took a step to use age, sex, and BMI as covariates in our statistical analysis to minimize

TABLE 3: Comparison of risk allele frequency of 28 SNPs tested for association with T2D in a Mongolian sample against 1000G populations.

Nearby gene	SNP	T2D risk allele	T2D risk allele frequency				(MGL – CEU)/CEU	F_{ST} with Mongolian sample			T2D association
			MGL ^a	CEU ^b	CHS ^c	CHB ^d		CEU	CHS	CHB	
<i>TMEM163</i>	rs6723108	T	0.98	0.51	1.00	1.00	0.94	0.66	0.01	0.01	Yes
<i>C6orf57</i>	rs1048886	G	0.21	0.14	0.06	0.10	0.55	0.01	0.07	0.04	No
<i>WFS1</i>	rs1801214	T	0.94	0.65	0.88	0.98	0.46	0.34	0.03	0.01	No
<i>BCL11A</i>	rs243021	A	0.70	0.48	0.66	0.63	0.45	0.06	0.23	0.19	Yes
<i>MTNR1B</i>	rs1387153	T	0.36	0.25	0.45	0.41	0.42	0.02	0.02	0.00	Yes
<i>GRK5</i>	rs10886471	C	0.77	0.55	0.76	0.79	0.42	0.21	0.00	0.00	No
<i>GRB14</i>	rs3923113	A	0.81	0.59	0.89	0.81	0.37	0.12	0.02	0.00	No
<i>RASGRP1</i>	rs7403531	T	0.36	0.28	0.29	0.34	0.29	0.01	0.01	0.00	No
<i>PRC1</i>	rs8042680	A	0.92	0.72	1.00	1.00	0.28	0.67	0.05	0.04	Yes
<i>IRS1</i>	rs7578326	A	0.79	0.66	0.83	0.86	0.20	0.04	0.00	0.01	Yes
<i>THADA</i>	rs7578597	T	0.98	0.88	0.99	1.00	0.12	0.13	0.00	0.00	Yes
<i>PAX4</i>	rs10229583	G	0.82	0.75	0.84	0.84	0.09	0.01	0.00	0.00	Yes
<i>CENTD2</i>	rs1552224	A	0.89	0.88	0.91	0.90	0.01	0.00	0.00	0.00	No
<i>HMG2</i>	rs1531343	C	0.11	0.11	0.12	0.11	0.00	0.00	0.00	0.00	Yes
<i>CDKN2</i>	rs1333051	A	0.83	0.84	0.88	0.86	–0.01	0.00	0.00	0.00	Yes
<i>SPRY2</i>	rs1359790	G	0.68	0.73	0.70	0.67	–0.07	0.00	0.00	0.00	No
<i>LGR5</i>	rs7961581	C	0.24	0.26	0.21	0.19	–0.09	0.00	0.00	0.00	No
<i>TP53INP1</i>	rs896854	T	0.39	0.43	0.30	0.31	–0.09	0.00	0.02	0.01	No
<i>IGF2BP2</i>	rs4402960	T	0.27	0.31	0.23	0.25	–0.11	0.00	0.00	0.00	No
<i>KCNJ11</i>	rs5215	C	0.34	0.38	0.40	0.38	–0.12	0.00	0.00	0.00	No
<i>KCNQ1</i>	rs163182	C	0.36	0.41	0.31	0.39	–0.13	0.00	0.00	0.00	Yes
<i>ADAMTS9</i>	rs4607103	C	0.57	0.78	0.64	0.57	–0.28	0.09	0.01	0.00	No
<i>ZFAND6</i>	rs11634397	G	0.23	0.34	0.07	0.08	–0.33	0.33	0.07	0.06	No
<i>KCNQ1</i>	rs2237897	C	0.58	0.94	0.66	0.62	–0.38	0.23	0.01	0.00	Yes
<i>FTO</i>	rs8050136	A	0.17	0.44	0.14	0.15	–0.63	0.19	0.00	0.00	No
<i>ZBED3</i>	rs4457053	G	0.09	0.30	0.04	0.07	–0.70	0.17	0.01	0.00	No
<i>TCF7L2</i>	rs7903146	T	0.06	0.31	0.03	0.02	–0.81	0.27	0.01	0.01	No

^a Mongolian sample of this study. ^{b,c,d} Population samples of Caucasian, Chinese in Beijing, and Chinese in Shanghai from the 1000G project, respectively.

the effect of this disparity. The follow-up larger scale study should recruit more matching control individuals to cases.

None of the SNPs we tested for association with T2D here has been previously implicated to be associated with lipids based on the NHGRI GWAS catalog (available at <http://www.genome.gov/gwastudies/>, November, 2012). However, we observed that association of T2D risk allele (A, major allele) of rs7578326 (*IRS1*) is associated with the increased level of TG. An elevated level of TG has been implicated as a risk factor of T2D, which likely resulted from the diminished activity of insulin causing inhibition of microsomal TG transfer protein activity [31]. The major allele (A) of a SNP in the upstream region of the same gene (*IRS1*), namely, rs2972146, is reported to be associated with an elevated level of TG as well [32]. rs297214 has nominal linkage disequilibrium with the SNP reported here ($r^2 = 0.3753$, 1000 G phase I), indicating a possibility that the T2D associated SNP or its proxy could be playing a role in pathogenesis of T2D or its related TG metabolism through *IRS1* activity. Further functional work will help to understand

the role of this SNP. Since we did not have information on previous medical history of treatment for either high cholesterol or T2D for the patients, it is possible that such treatments, if administered previously, could have prevented us from seeing the effect of SNP association with the diabetes related quantitative traits, including TG.

Mongolians are one of the people who reside on the Mongolian Plateau in Asia and have heavily depended on nomadic life styles with harsh environments characterized by low temperature and scarce availability of food sources [33]. Although we do not have any concrete evidence that these variants might play a role in conserving energy, we found six T2D associated SNPs in our sample that have 10% or higher risk allele frequency compared to Caucasian samples. These SNPs have a comparable risk allele frequency with Chinese populations in the 1000 G project, indicating that our allele frequency estimate is reflecting the allelic structure of the SNPs in the populations of Asia. We observed that SNPs rs6723108 (*TMEM163*) and rs8042680 (*PRC1*) have substantial allelic differences in our sample compared

to Caucasians and the risk alleles have reached near fixed level in Mongolians. On the other hand, a widely replicated T2D SNP, rs7903146 (*TCF7L2*), in European populations has substantially low risk allele frequency in the Mongolian sample (0.06 versus 0.31 in the Caucasian sample; $F_{ST} = 0.27$) and is not associated in our sample of modest size. It is likely that the frequency difference of some SNPs between our sample and European population also contributed to the lack of reproducibility in T2D association in our study.

Although we note that rs6723108 has a wide range of OR estimate (95% CI 2–33.3; risk allele frequency 0.98), the OR of 7.7 in our study is substantially greater than what was reported in the original study in an Indian population [23] (OR = 1.31, 95% CI 1.20–1.44; risk allele frequency 0.89), indicating the potential higher effect of the variant in our sample. More systematic studies of population specific allelic architecture with respect to T2D are needed to dissect the potential impact of these highly differentiated SNPs in different populations.

In conclusion, our association study has confirmed association of several previously identified T2D susceptibility loci in the Mongolian sample. We also identified rs7578326 near *IRS1* to be associated with an increased level of TG. The observation of the remarkable allele frequency difference of the T2D SNPs in our sample compared to Caucasians is important in further identifying causative variants for T2D and understanding the role of these SNPs in development of T2D in different ethnic populations.

Conflict of Interests

The authors declare that they have no competing interests with respect to the present paper.

Authors' Contribution

Haihua Bai, Qizhu Wu, and Narisu Narisu conceived and designed the study. Haihua Bai, Suyalatu Suyalatu, Shandan Chu, and Burenbatu Borjigin collected the samples. Qizhu Wu, conducted clinical diagnoses, performed experiments, interpreted the data, and obtained funding. Haihua Bai and Haiping Liu prepared DNA. Narisu Narisu, Haihua Bai, Haiping Liu, and Xiaosen Guo analyzed and interpreted the data and drafted the manuscript. Xiaosen Guo, Yuriy L. Orlov, and Tianming Lan prepared the sequencing libraries and performed genotyping and calling variants. All other authors provided technical assistance and interpreted the data. All authors read and approved the final paper. Haihua Bai, Haiping Liu, and Suyalatu Suyalatu contributed equally to this work.

Acknowledgments

The authors thank their study individuals for their generous participation in this study. This study was supported by the National Natural Science Foundation of China (81160101 and 81060098) and NSFC-RFBR Cooperation and Exchange Projects (81511130050). Narisu Narisu was supported by the intramural program of the National Human Genome

Research Institute, National Institutes of Health. Yuriy L. Orlov, Olga L. Posukh, and Ludmila P. Osipova were supported by RFBR (15-54-53091). The authors thank Anne Jackson and Clement Ma for advice in statistical analysis and comments in writing. They also thank Francis S. Collins for critical comments.

References

- [1] W. Yang, J. Lu, J. Weng et al., "Prevalence of diabetes among men and women in China," *The New England Journal of Medicine*, vol. 362, no. 12, pp. 1090–1101, 2010.
- [2] L. A. Hindorf, P. Sethupathy, H. A. Junkins et al., "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 23, pp. 9362–9367, 2009.
- [3] The GoT2D Consortium, "Whole-genome sequencing of 2,657 individuals demonstrates the primary role of common variation in genetic risk of type 2 diabetes," In press.
- [4] J. M. Keaton, J. N. Cooke Bailey, N. D. Palmer et al., "A comparison of type 2 diabetes risk allele load between African Americans and European Americans," *Human Genetics*, vol. 133, no. 12, pp. 1487–1495, 2014.
- [5] X. Sim, R. T.-H. Ong, C. Suo et al., "Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia," *PLoS Genetics*, vol. 7, no. 4, Article ID e1001363, 2011.
- [6] A. Mahajan, M. J. Go, W. Zhang et al., "Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility," *Nature Genetics*, vol. 46, no. 3, pp. 234–244, 2014.
- [7] H. Bai, X. Guo, D. Zhang et al., "The genome of a Mongolian individual reveals the genetic imprints of Mongolians on modern human populations," *Genome Biology and Evolution*, 2014.
- [8] Z. Odgerel, H. S. Lee, N. Erdenebileg et al., "Genetic variants in potassium channels are associated with type 2 diabetes in a Mongolian population," *Journal of Diabetes*, vol. 4, no. 3, pp. 238–342, 2012.
- [9] B. Burenbatu, Z. Li, and Y. Wang, "A report of chronic diseases in Mongolian, Chinese and other minorities in Tongliao city," *Zhong Guo Man Xing Bing Yu Fang Yu Kong Zhi*, vol. 20, no. 1, pp. 83–85, 2012.
- [10] L. Haiping, L. Yuehui, M. Bangzhu et al., "The study of diabetes in Mongolians in China," *Journal of Inner Mongolia Normal University*, vol. 29, no. 4, pp. 401–403, 2014.
- [11] Diabetes Consortium of China, "Report of diabetes in 300,000 people of 14 provinces of China," *Zhong Hua Nei Ke Za Zhi*, vol. 20, no. 11, p. 678, 1981.
- [12] WHO, *Screening for Type 2 Diabetes*, WHO, 2003.
- [13] B. Cui, X. Zhu, M. Xu et al., "A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese," *PLoS ONE*, vol. 6, no. 7, Article ID e22353, 2011.
- [14] H. Unoki, A. Takahashi, T. Kawaguchi et al., "SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations," *Nature Genetics*, vol. 40, no. 9, pp. 1098–1102, 2008.
- [15] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.
- [16] H. Li, B. Handsaker, A. Wysoker et al., "1000 Genome project data processing subgroup. The sequence alignment/map format

- and SAMtools,” *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.
- [17] H. M. Kang, “EPACTS: efficient and parallelizable association container toolbox,” 2014, <http://www.sph.umich.edu/csg/kang/epacts/>.
- [18] G. R. Abecasis, A. Auton, L. D. Brooks et al., “with 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes,” *Nature*, vol. 491, no. 7422, pp. 56–65, 2012.
- [19] S. Purcell, B. Neale, K. Todd-Brown et al., “PLINK: a tool set for whole-genome association and population-based linkage analyses,” *The American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–575, 2007.
- [20] B. S. Weir and C. C. Cockerham, “Estimating F-statistics for the analysis of population structure,” *Evolution*, vol. 38, no. 6, pp. 1358–1370, 1984.
- [21] A. D. Mooradian, “Dyslipidemia in type 2 diabetes mellitus,” *Nature Clinical Practice Endocrinology & Metabolism*, vol. 5, no. 3, pp. 150–159, 2009.
- [22] Z. Odgerel, H. S. Lee, N. Erdenebileg et al., “Genetic variants in potassium channels are associated with type 2 diabetes in a Mongolian population,” *Journal of Diabetes*, vol. 4, no. 3, pp. 238–242, 2012.
- [23] R. Venkatesan, V. Mohan, N. Tandon et al., “Genome-wide association study for type 2 diabetes in indians identifies a new susceptibility locus at 2q21,” *Diabetes*, vol. 62, no. 3, pp. 977–986, 2013.
- [24] R. C. W. Ma, C. Hu, C. H. Tam et al., “Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near *PAX4*,” *Diabetologia*, vol. 56, no. 6, pp. 1291–1305, 2013.
- [25] E. Zeggini, L. J. Scott, R. Saxena et al., “Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes,” *Nature Genetics*, vol. 40, no. 5, pp. 638–645, 2008.
- [26] B. F. Voight, L. J. Scott, V. Steinthorsdottir et al., “Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis,” *Nature Genetics*, vol. 42, no. 7, pp. 579–589, 2010.
- [27] E. J. Parra, J. E. Below, S. Krithika et al., “Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas,” *Diabetologia*, vol. 54, no. 8, pp. 2038–2046, 2011.
- [28] F. Takeuchi, M. Serizawa, K. Yamamoto et al., “Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population,” *Diabetes*, vol. 58, no. 7, pp. 1690–1699, 2009.
- [29] X. O. Shu, J. Long, Q. Cai et al., “Identification of new genetic risk variants for type 2 diabetes,” *PLoS Genetics*, vol. 6, no. 9, Article ID e1001127, 2010.
- [30] H. Li, W. Gan, L. Lu et al., “A genome-wide association study identifies *GRK5* and *RASGRP1* as type 2 diabetes loci in Chinese Hans,” *Diabetes*, vol. 62, no. 1, pp. 291–298, 2013.
- [31] R. Bitzur, H. Cohen, Y. Kamari, A. Shaish, and D. Harats, “Triglycerides and HDL cholesterol: stars or second leads in diabetes?” *Diabetes care*, vol. 32, Supplement 2, pp. S373–S377, 2009.
- [32] Global Lipids Genetics Consortium, “Discovery and refinement of loci associated with lipid levels,” *Nature Genetics*, vol. 45, no. 11, pp. 1274–1283, 2013.
- [33] L. Morris, “Death and suffering in the land of Genghis Khan,” *Canadian Medical Association Journal*, vol. 183, no. 5, pp. E271–E272, 2011.

